

Summary

Contents

A Quarterly Information Bulletin from the Methodology Division

Contents

Sampling Strategies for Indigenous Household Surveys
Emerging Areas of Analytical Expertise
Generalised Regression Estimation for ABS Business Surveys
Medical Survey Design
Developing and Managing Professional Expertise

Sampling Strategies for Indigenous Household Surveys

SAMPLING STRATEGIES FOR INDIGENOUS HOUSEHOLD SURVEYS

The expanded ABS household survey program involves a series of regular large scale surveys of Aboriginal & Torres Strait Islander people. Upcoming surveys include the Indigenous Social Survey (ISS) and an Indigenous Supplement to the 2004 National Health Survey (NHSI). These surveys both aim to enumerate in excess of 10,000 Aboriginal & Torres Strait Islander people Australia wide, a significant proportion of the Indigenous population. The surveys present numerous challenges to all aspects of the survey process, particularly the development of appropriate sampling methodologies.

Remote Area Sampling

Approximately 20% of the Indigenous population reside in remote areas as defined in the Access/Remoteness Index for Australia (ARIA). These areas are in-scope of Indigenous household surveys, unlike most other ABS special social household surveys. Surveying in remote areas presents many problems, including

- the very high cost of sending interviewers out to these areas;
- protocol in approaching and conducting interviews in Indigenous communities; and
- high levels of sample loss.

The 2001 NHSI saw the introduction of a list based frame into an ABS household survey which aimed to address the problems of sampling in remote areas. The frame was constructed from the 1999 ABS Community and Housing Infrastructure Needs Survey (CHINS), an administrative survey which collected information on housing and infrastructure from all Indigenous communities in Australia. In remote areas the Indigenous population resides largely in communities which are registered on CHINS. It was considered that the operational advantages of using a list based frame of CHINS communities far outweighed the impact of slight undercoverage. These advantages include:

- more precise identification of where Indigenous people reside thereby avoiding high costs of screening large remote Collection Districts (CDs) (see non-remote sampling below);
- appropriate formation of groups of associated communities, including smaller communities known as out-stations, into primary sampling units enabling interviews to be conducted more effectively with necessary permission and facilitation; and
- a more cost effective sample.

As part of the Monthly Population Survey (MPS) redesign an Indigenous community stratum is being developed. This stratum will form the basis of the 2002 ISS and 2004 NHSI remote area samples. Although sample selection within this stratum will be CD based, the general sample selection concepts underpinning the 2001 NHSI sparse sampling methodology will be adopted. In particular, the 2001 CHINS will be crucial to the formation of groupings of CDs which will form the stratum frame. In addition, the use of a CD based frame will facilitate overlap control between the MPS and Indigenous special social surveys.

Non-Remote Area Sampling

Sampling Indigenous people in non-remote areas faces the daunting prospect of attempting to sample a rare population whose exact location is unknown. Sampling in non-remote areas involves a large search phase to initially identify households which contain Indigenous persons. It is this facet of the sampling methodology which is the most problematic. Based on 1996 Census figures, approximately 7% of the Indigenous population reside in CDs which contain only one Indigenous household. For some states this figure is much higher, (e.g. 28% in Victoria). Due to the significant numbers of Indigenous people residing in low Indigenous household density CDs it is not feasible to justify putting them out of scope of Indigenous surveys. This invariably leads to very high screening costs in order to identify and enumerate Indigenous households in these areas. For example, the 2001 NHSI non-remote sample aimed to select approximately 1065 fully responding Indigenous households. In order to achieve this, roughly 37,250 households were selected to be screened!

Adding to the problem is the mobility of the Indigenous population. For example, the 2001 NHSI sample design was based on Census figures which were 5 years out of date. This has resulted in lower than expected Indigenous sample takes in some areas, leading to a less efficient than expected sample.

For the upcoming 2002 ISS a number of methodologies are being investigated to develop more effective and efficient sampling methodologies for non-remote Indigenous surveys. These include:

- stockpiling households from outgoing MPS rotation groups identified as containing Indigenous persons and then using these households to represent low Indigenous household density CDs;
- putting low density Indigenous household CDs out of scope in some states;
- introducing quota sampling for some strata in order to more effectively control the number of households screened;
- sampling next door neighbours of identified Indigenous households in order to maximise the likelihood of obtaining expected Indigenous household sample size in the field.

Suggestions on alternative screening mechanisms are most welcome!

For more information please contact Alistair Rogers on (02) 6252 7334

Email: al.rogers@abs.gov.au

Emerging Areas of Analytical Expertise

EMERGING AREAS OF ANALYTICAL EXPERTISE

As well as maintaining our traditional skills, Analysis Branch must acquire expertise in new methods - both methods that have recently emerged in the literature and some not-so-recent methods that we have not before applied to ABS work.

Thinking about our needs for new expertise has been prompted by several developments:

- increasingly sophisticated demands from our clients (and concerns about ensuring that our work is professionally defensible);
- changes to the kinds of data we are analysing (including by-products of administrative or business processes);
- rethinking some aspects of statistical quality (and ensuring that the quality and other characteristics of analytical products are visible to users); and
- changes to the kinds of work our ABS colleagues are undertaking and the kinds of skills they wish to acquire (in line with Statistical Skills of ABS Staff and other key documents).

The areas in which we might acquire expertise are very diverse. Some possibilities include new methods for:

- producing estimates for small domains (e.g. small geographical areas, subpopulations or subindustries);
- taking account of complex survey designs (clustering, multistage selection and the like);
- taking account of the multilevel character of some socioeconomic phenomena (e.g. individual, group and area influences on crime)
- modelling with large, dirty, melded datasets; and
- doing longitudinal and dynamic analyses on datasets that have not been collected through a truly longitudinal survey.

Our prime criterion for choosing which areas we shall invest in is "How broadly can a new method be applied to core ABS concerns and how large a benefit might the method deliver?"

We are also thinking about how we can gather and spread intelligence about emerging methods. One scheme we are experimenting with is to appoint "gatekeepers" whose role is to trawl the literature in a given field, (see article on Developing and Managing Expertise).

For more information about emerging areas of analysis, please contact Ken Tallis on (02) 6252 7290.

Email: ken.tallis@abs.gov.au

Generalised Regression Estimation for ABS Business Surveys

GENERALISED REGRESSION ESTIMATION FOR ABS BUSINESS SURVEYS

The availability of Business Activity Statement (BAS) data collected by the Australian Taxation Office provides the ABS with opportunities to improve the efficiency of sample design and estimation for its business surveys. The ABS business surveys currently use two methods of estimation; number-raised and ratio. While ratio estimation allows the use of a single auxiliary variable to improve the precision of the estimates, generalised regression (GREG) estimation allows the use of more than one auxiliary variable, and hence has the potential to be more efficient (i.e. reduce the current sample sizes for the ABS business surveys with no reduction in the accuracy of the estimates) than number-raised and ratio estimation.

The GREG development project will be conducted in four distinct phases.

Phase 1

The first phase was to conduct an initial evaluation of whether GREG estimation has the potential to deliver any reductions in the current sample sizes for the ABS business surveys. An investigation conducted by Methodology Division (MD) on several business surveys found that using GREG estimation with BAS data should result in substantial reductions in the current sample sizes for ABS business surveys.

Phase 2

The second phase is to further develop and refine the generalised regression estimation techniques, and develop a number of 'environment independent' components that will calculate generalised regression estimates and accuracy measures of these estimates (referred to as the GregEst components). These can then be incorporated into existing ABS business processes. At this stage, MD have almost completed the development of methodological techniques for weighting, estimation, winsorization and variance estimation (using a bootstrap replication method) under a generalised regression estimation framework, while the Technical Services Division have commenced the design and construction of the GREG components to perform these methodological techniques.

Phase 3

The third phase is to evaluate the GREG estimation methodology using BAS data and the GregEst components on a number of business surveys, with the aim of developing business cases for using BAS data in estimation. This phase is expected to commence after the construction and testing of the GREG components has been completed.

Phase 4

The fourth phase is to implement the GREG methodology and support the GregEst components into business processes.

For more information, please contact John Preston on (02) 6252 6970.

Medical Survey Design

MEDICAL SURVEY DESIGN

The 2001/2 Medical Survey is despatched in two stages and is used to produce two publications. Practitioners are the statistical unit for the first stage of the survey while the medical and administrative services businesses for which the practitioners work form the statistical unit for the second stage despatch.

The first stage involves selection of a random sample of practitioners from a list of practitioners obtained from the Department of Health and Aged Care's Medicare Provider File. The first stage despatch, which occurred in February, asks the selected practitioners to provide details about themselves (such as hours worked, main field of work, number of private patient contacts in an average week) as well as details of the medical practices they work for (such as name, address, ABN, main income earning activity of business).

A second stage despatch will occur in August 2002 and will consist of the selection of all in-scope practices identified by the practitioners selected in the first stage despatch. In addition, a small number of pathology labs will be selected from a frame created from a number of sources. The second stage despatch will seek detailed financial and employment data from the medical practices. Business details collected from practitioners in the first stage despatch will be used in the identification and removal of duplicate practices prior to the second stage being despatched.

The Medicare Provider file is used as the frame for the first stage survey instead of the ABS Business Register because:

- The ABS Business Register does not have a list of individual practitioners;
- The general structure of medical 'practices' is a non-standard unit concept. 'Practices' are formed from the combined operations of medical businesses and their associated administrative services businesses, with the latter generally classified to non-medical industry classifications (i.e. ANZSICs). The practice represents the medical 'industry' unit of interest for users.
- The ABS Business Register does not allow for the stratification of specialists into the required individual specialties (eg. anaesthesia, paediatrics, dermatology, pathology, surgery, psychiatry).

A survey of the private medical practice industry was previously conducted in respect of 1994-95. A few changes have occurred since the last survey including:

- the requirement to output estimates for General Practitioners by RRMA (Remote, Rural and Metropolitan Areas classification).
- the recent emergence of corporates in the private medical practice industry (i.e. non-medical businesses managing private medical practices).

The first stage sample is a stratified simple random sample. In the second stage, all practices identified are completely enumerated. It should be noted however that the sample design doesn't strictly meet the definition of a two stage design because the first stage sampling units (practitioners) are a subset of the second stage sampling units (practices). Also, conventional two stage formulae don't hold due to the fact that it is possible to have a doctor work in several practices, and a practice can have several doctors working in it.

The 94/95 Medical survey dataset was used as design data in the sample design. Optimal allocation was used to determine the sample sizes required for both stages of the survey, taking account of separate constraints and design variables for the two stages.

The Stage 1 despatch will be sent to 3,100 practitioners from of a population of over 37,000. The second stage despatch is expected to be sent to approximately 4,000 businesses.

For more information please contact Adam Thomas on (03) 9615 7406

Email: adam.thomas@abs.gov.au

Developing and Managing Professional Expertise

DEVELOPING AND MANAGING PROFESSIONAL EXPERTISE

This article reports on the activities of the Breadcrumbs Nuggets and Roadmaps (BNR) team. The BNR team's objective was to generate ideas, discussion and debate within the Methodology Division (MD), with a view to improving the way MD develops and manages individual and collective professional expertise.

Before discussing the activities of the team, it is instructive to reflect on the title for the group and the metaphors contained within.

A breadcrumb is a little piece of data or knowledge. When you are working on a project you find (and sometimes create) them scattered all around, sometimes organised usefully, but more often in bits and pieces. A key research skill is to find relevant breadcrumbs and organise them to form a useful body of knowledge. A good trail of breadcrumbs can lead you safely through the project to the end.

A nugget is a piece of treasure that you find, when someone before you has gathered all the breadcrumbs together, and drawn connections between ideas, techniques or whatever that originally might have seemed quite unrelated. Depending on what you are interested in, some breadcrumbs might be part of several different nuggets simultaneously.

A roadmap is something really valuable. It is something that people who have gone before sometimes leave for you - a picture of how to find your way along trails of breadcrumbs, what nuggets you can find along the way, and what is the quickest way to reach them. Roadmaps aren't always linear - sometimes there are different ways to get to the same place, and conversely some nuggets are on the paths to several different destinations!

The BNR team decided to pursue a number of 'focused' activities as well discussing issues in developing and maintaining professional expertise.

Rob Burnside considered better ways to document and record research and information produced in the course of work. This was based initially on interviews with ABS staff, and aimed at uncovering information storage and retrieval best practices by looking at how people actually undertake their own research and retrieval. Three interesting threads emerged;

- Information retrieval patterns and behaviour often follow from professional skills, training, monitoring of journals, mailing lists and databases and copying articles.
- The almost unconsciously accumulated awareness of the interests and skills of other individuals. Lack of this awareness is a problem for people new to an organisation or work area.
- The desire, need and means of attributing some level of reliability to information uncovered, and the way to make informal assessments of sources based on a range of factors. These include the degree to which papers are circulated, associated seminars and minutes, the quality of writing, and status of the source area or author.

An implication of this is that authors and managers of knowledge should aim to explicitly support and facilitate accurate and effective warranting by readers. For example avoid leaving documents with titles that include 'final draft' or 'interim' and include clear cues as to role, responsibility and organisational information.

Craig McLaren and Sybille McKeown examined how information was stored on MD's large and expanding databases. While the current organisation of databases was useful for finding project or client specific information, it was not as effective for finding technique or method-specific information. To address this deficiency they developed a list of keywords that can be attached to important or final documents and used to search by. A set of keywords are currently being trialled on the MD databases.

Steven Kennedy with the help of Kristen Northwood and Godfrey Lubulwa explored the various methods people use to collect and disseminate the knowledge they have gained on trips and at conferences. Seminars were usually the most useful way of transmitting what people had learned to others. Also, thematic seminars as opposed to descriptions of trips were a good choice when the traveller had attended a conference or workshop. Producing a report was a useful way of clarifying and transmitting what people had learned from their travel. It appears that using a dictaphone daily while travelling to record comments and then asking for these comments to be typed up once home again was the most efficient mechanism for producing a record of the trip, from which the final report could be developed. Even so substantial effort is still needed to appropriately structure the final report.

Ken Tallis concentrated on the knowledge that arises from projects. The key question was 'What tools and practices would help ensure that such knowledge is captured, shared and remains accessible to future researchers in MD

and elsewhere?' Recommendations for standard outputs from methodological projects included roadmaps to key project documents, annotated bibliographies, and narratives about research strategies (especially, what strategies worked and what didn't). Other useful tools include roadmaps of analytical techniques and software, and guides to the expertise of ABS methodologists.

Geoff Lee looked at annotated bibliographies as a method for telling others about interesting and useful papers. For each journal article, he made an entry on our Lotus Notes database which included an 'opinion'. The 'opinion' reflected what was learnt from or thought of the paper and (hopefully) added value for later researchers who came across this 'breadcrumb'. The second part of the project investigated some tools to help assemble individual library entries and opinions into a completed reading list (a nugget). A Notes folder was created, into which each library entry could be placed. The entries could be sorted into author order or chronological order and an annotated reading list (the final nugget) could be automatically generated. The annotated reading list contained the title and author of each paper, followed by a link to the corresponding library form and the 'opinion', so later readers could decide if the paper is relevant to their research interest. While not yet implemented, standard keywords would also be recorded on the annotated reading list documents.

Tala Talagaswatta and Jonathon Khoo are considering how to gather and spread intelligence about emerging analytical methods. One scheme they are experimenting with is 'gatekeepers'. The gatekeeper's role is to trawl the literature in a given field, subscribe to relevant internet sites or groups, encapsulate the emerging methods, and then bring and spread that knowledge to the Methodology Division. A trial is currently running in the Analytical Services Branch.

A poster will be distributed shortly that summarises all the BNR activities to date. People who are interested in any aspect of the BNR team's investigations should contact Geoff Lee on (02) 6252 5239

Email: geoff.lee@abs.gov.au.

About this release

Methodological News

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence